

Seven Food Stories from Seven Cities^{*}

Jing Ouyang¹ Tong Wu¹ Jing Zhou¹

University of Southern California, Los Angeles CA 90007, USA

Abstract. In this paper, we use Yelp dataset to analyze the restaurant data and present recommendations for users with visual design and use HTML, D3, ArcMap and other tools to illustrate our points. We get inspirations from previous works and improve their designs with the method we learned from class.

Keywords: Yelp · restaurant · data visualization.

1 Introduction

With the increasing number of immigration and international visitors, the United States and Canada have more diversity on their food genre. People would love to enjoy not only traditional and local food, but also desire to explore different type of food with the help of recommendation applications, such as Yelp. For this final project, we would love to use Yelps data for analysis which provides information and suggestions for users. In order to achieve this goal, we would enhance data visualization methods such as graphs, maps and charts to illustrate our points, then we make webpages to present our result for users. We would talk about this project from four different aspects, including literature review, story introduction, graph design and conclusion.

2 Literature Review

Before we narrow our scope down to the Yelp dataset, we searched several datasets online for previous works on relative topics. There were several data analysis case studies and data visualization on Yelp dataset [1]. For the majority of studies, they used the whole dataset which included restaurant rating, sight-seeing rating, shopping center rating, etc. other than that, they applied ggplot from R to visualize the dataset with bubble plot, box plot, scatter plot and mapplot [2]. However, they were only present the distribution of location and the relationship between rating and location. Some other studies were focusing on review analysis over time, such as review trend over time for one specific restaurant or a category of restaurant [3].

Other than that, we found out that several scholars would like to use pie chart and bar chart to highlight the distribution of the content of our target dataset.

^{*} Team Jings.

Additionally, they made some improvements to those charts by reordered the categories and changed the color combination to enhance the contrast. They also used word cloud to analysis the review in order to get a summary for target restaurant [4]. Unfortunately, the result was neither pleasant nor efficient since there were abundant reviews about services, locations and reservations rather than the food. Therefore, users could not get the information they wanted through the word cloud.

Except creating charts and graphs, previous researchers also used maps to offer a straight forward way to illustrate the feature of this dataset. By doing the literature review, we had a draft about how to apply those data visualization methods on our target dataset, combine information with users needs and avoid problems while making our own websites. The main problem of all previous work was the dataset, therefore, we decided to use only the restaurant data from Yelp dataset and focus on five cities which included Madison, Las Vegas, Phoenix, Cleveland, Charlotte, Montreal and Toronto.

3 Story Introduction

Our final project is aiming to present seven stories about food depends on seven different cities. We would love to show the feature of food in different cities and make recommendations for our users based on Yelp dataset. First, we would like to compare these seven cities in a general level. Therefore, our users could see the big picture of those seven cities food industry. Then, we would be willing to elaborate the story of each city based on their features, such as their traditional food, trending food, top rating food genres and recommendations. For this part of project, we plan to do some research about the most famous and iconic food for each city and visualize it with flat design and catching design. After that, we would provide top seven recommendations based on Yelp rating and attach brief introductions of each restaurant.

In order to show our innovation, we generate subtitles for each city and use bar charts, graphs and pictures to illustrate our points. For example, we name Las Vegas as the city never sleep and we use a time slider map animation to show the opening and closing time of restaurants in Las Vegas and compare it with other six cities to prove our points.

After analysis the data, we realize that different city has significant aspects. For example, Toronto is a diverse city which has the most type of restaurant among our target cities. Also, Toronto has many Asian immigrations, therefore, we offer more information of Asian food and provide the ranking of Asian food. In order to provide an overall view of Toronto, we would like to show the average review counts per restaurant and the frequency of restaurant types with two bar charts.

Nowadays, people have the trend to try various type of food while enjoy the most authentic local cuisine, therefore, users might want to try something uncommon. So, we argue that presenting the high rating uncommon restaurant would match their needs. We also want to present the top high-ranking restau-

rants, lowest ranking restaurants and high-ranking iconic restaurants. In order to visualize this part, we need to filter out the minority food genre in each city and then get the most popular and high rated restaurants. Finally, we could show both of the iconic food and uncommon food at the same time.

4 Graph Design

In order to tell the different stories of those seven cities, we import geojson files from carto [5] for Las Vegas and from Github [6] for the rest of those cities. At first, we want to use the geographic map for this part, however, we prefer the simplicity style. So, we use white background and lines to represent different parts of those cities. Then we apply dot map to show the distribution of restaurants while add zoom in effect for all of them. Users could drag the map to see distinctive areas and the dots keep at the same level of size and relative locations. Next, we realize when we zoom in the map, the size of dot might be disturbing. Therefore, we implement d3 [7] to change the dot size relative to the map scale. This method is adopted from our lecture and it is a user-friendly tool for many maps and graphs. In order to show more information while keeping the map clean, users would see the more detail information of the restaurant such as address, introduction and services time when they hover the dot.

During this project, we would like to compare the features of restaurant among those seven cities, so we have to create various bar charts and stack charts. Therefore, we use a function in d3 to create those charts instead of making them one by one with hard code methods. We engage user interaction with bar charts and stack charts since our users could choose their favorite places, food genres and restaurant types.

Furthermore, we employ coxcomb chart to show the comparison of several features of different restaurants at the same time with the online template [8]. With the coxcomb chart, our users are easily navigating the interface and see the difference among different cities. Users could also see the frequency of rating for several type of restaurants while show this data in different cities. Since the chart has to get at least two input, we design to lock the check box when there are only two features checked.

Finally, we use bootstrap [9] to design the whole webpage since it is not only easy to navigate, but also suitable for our topic. It could give us the simple and clear home page and variously designed subpages.

5 Conclusion

We started from another idea which was also interesting, however, we could not find enough data and previous researches to work with. Therefore, we changed our topic several times to adjust and tried to reach a balance between our interest and the reality. This final project did not only help us to apply all of the data visualization method that we learned from lecture, practice our website design coding skills outside of our required assignments. The most important thing

about this final project was pushing us to learn by ourselves and trained us to get used to solve unexpected problems while making websites. Those problems and bugs were hard to forecast while planning, therefore, we had to get use to solve problems and change our original plan while still achieve our goal.

The whole team had many brilliant ideas while brain storming and everyone knew themselves better while they were finishing the final project. We all contributed to this project with our efforts and thought as a team to work together. If we have more time and skills, we would polish our project and make more content from the dataset.

References

1. Kaggle Yelp dataset page, <https://nycdatascience.com/blog/student-works/project-1-exploratory-visualizations-of-yelp-academic-dataset-draft/>. Last accessed 25 Nov 2018
2. Kaggle Yelp dataset page, <https://www.kaggle.com/vksbhandary/exploring-yelp-reviews-dataset>. Last accessed 20 Nov 2018
3. Kaggle Yelp dataset page, <https://www.kaggle.com/ibadia/visualization-on-yelp-dataset>. Last accessed 24 Nov 2018
4. Kaggle Yelp dataset page, <https://www.kaggle.com/ambarish/a-very-extensive-data-analysis-of-yelp>. Last accessed 22 Nov 2018
5. Carto Las Vegas geojson page, <https://carto.com/>. Last accessed 25 Nov 2018
6. Carto Las Vegas geojson page, https://github.com/codeforamerica/click_that_hood/. Last accessed 26 Nov 2018
7. d3, <https://d3js.org/d3.v5.min.js>. Last accessed 24 Nov 2018
8. Coxcomb chart template, <https://bl.ocks.org/mbostock/6fead6d1378d6df5ae77bb6a719afcb2>. Last accessed 23 Nov 2018
9. Bootstrap template, <https://startbootstrap.com/template-overviews/agency/>. Last accessed 26 Nov 2018